

October 29, 2024

Hon. Ona T. Wang  
United States Magistrate Judge  
Southern District of New York  
New York, New York 10007

Re: *The New York Times Company v. Microsoft Corporation, et al.*, Case No.: 23-cv-11195-SHS-OTW and *Daily News LP, et al. v. Microsoft Corporation, et al.*, Case No.: 24-cv-03285-SHS-OTW: Discovery Dispute Regarding OpenAI's Responses to The Times's Second and Third RFPs

Dear Magistrate Judge Wang:

The Times again burdens this Court and surprises OpenAI with a premature motion. OpenAI has been conferring with The Times in good faith and agreed to produce documents at the core of the RFPs referenced below: Dnyuz, web crawlers, and the removal of copyright management information from training datasets. OpenAI has continued to confer, however, about a reasonable search methodology and scope related to third-party websites, third-party web crawlers, and the technology known as retrieval augmented generation (which is not covered by The Times' referenced RFP). Rather than continuing to meet and confer in good faith with OpenAI, as required by this Court's orders, The Times is jamming an already-packed status conference agenda with unripe issues. The Court should deny The Times's motion and order the parties to continue conferring.

### **1. RFPs 84 and 85**

There is no dispute as to RFP 84. OpenAI told The Times that it has already run "DNYuz" as a search term and has been reviewing the resulting hits. Dkt. No. 262 at 34. Contrary to The Times's representations, OpenAI has never suggested otherwise. Dkt. No. 283 at 2.<sup>1</sup> In any event, this is a non-issue: OpenAI has now produced some DNYuz documents, and OpenAI will continue to review and produce responsive, non-privileged documents. The Times would have known this if it had heeded this Court's clear directive to continue to meet and confer at the first signs of disagreement. *See* 9/12/2024 Hearing Tr. 22:15–22. Instead, The Times filed its motion without warning. Given OpenAI's production and ongoing review, this portion of The Times's motion is moot.

As to RFP 85, The Times again failed to meet and confer in good faith. The Times is correct that OpenAI "claims this RFP is overbroad"—which is precisely why OpenAI asked The Times to confer about The Times's understanding of unspecified "websites besides DNYuz." *See* Dkt. No. 262 at 34; Dkt. No. 283-7 at 11. Just days ago, The Times told OpenAI that it would "write separately regarding RFP 85." *Id.* at 8.

---

<sup>1</sup> When The Times demanded a hit report for the term "dnyuz," OpenAI simply told The Times to follow the procedure for obtaining such reports set forth in the ESI protocol. Dkt. No. 287-3 at 8; Dkt. No. 251 ¶¶ 4-5.

But it never did. Instead, it filed this motion, which *for the first time* offered to narrow the request to documents about “websites . . . known to pirate publisher content” and “relating to copying and plagiarism.” Dkt. No. 283 at 2. OpenAI is willing to search for such websites, if The Times will provide a list of their names. Accordingly, The Times’s motion should be denied, and the parties should continue to meet and confer regarding the scope of RFP 85.

The Times’s failure to meet and confer is inefficient and unproductive. OpenAI first learned that the parties were supposedly “unable to resolve [these] RFPs” from The Times’s motion. Dkt. No. 283 at 1. The parties had met and conferred *just eight hours before* The Times’s filing. Nikolaides Decl. ¶ 2. During that conference, The Times did not raise these RFPs, nor did it provide any other indication that it considered the parties to be at an impasse on these requests or would be moving shortly. *Id.* ¶¶ 3-4. OpenAI has not “refused to produce any documents in response to RFPs 84 and 85,” as The Times contends. Dkt. No. 283 at 1. There has been no “about-face”—OpenAI has now produced documents responsive to RFP 84, and as OpenAI explained in the parties’ joint agenda for the upcoming status conference, it “remains eager to meet and confer to better understand The Times’s position on these requests.” Dkt. No. 262 at 35.

## 2. RFP 26

The parties were making significant progress regarding the scope of RFP 26 in the days leading up to The Times’s filing, which makes The Times’s motion puzzling. OpenAI had agreed to conduct a reasonable search for and produce documents concerning OpenAI’s use of web crawlers to crawl content of articles published by The Times in connection with training the relevant models. Dkt. No. 262 at 30; Dkt. No. 283-7 at 4. As The Times notes, this narrowed the issues to only one dispute: whether OpenAI must *also* produce documents regarding the use of web crawlers by third parties that OpenAI does not control.

Just days ago, OpenAI reiterated that it has no control over nor insight into how these third parties use web crawlers. *Id.* But OpenAI remained committed to meeting and conferring.

Instead, The Times filed this motion without warning. And now, for the first time, The Times cites a set of “[d]ocuments produced by Microsoft.” Dkt. No. 283 at 2. The Times never shared these documents with OpenAI despite repeated opportunities to do so. Moreover, these documents do not support The Times’s request for OpenAI to perform a search for documents about third parties OpenAI does not control. These *Microsoft* documents prove OpenAI’s point: OpenAI has limited knowledge about Microsoft’s or any other third party’s use of web crawlers. As with RFPs 84 and 85, The Times’s request is premature and should be denied. The more efficient approach is for the parties to continue meeting and conferring regarding a mutually acceptable scope for RFP 26.

## 3. RFP 18

RFP 18 seeks documents concerning the alleged removal of copyright management information (“CMI”) from OpenAI’s training datasets, *i.e.*, during the training process. After negotiating,

OpenAI thought that the parties had reached an agreement on the scope of this RFP: non-privileged documents located after a reasonably diligent search concerning alleged removal of CMI *from training datasets*. Dkt. No. 283-7 at 4. But just two days before The Times filed its motion, it suddenly claimed that this RFP also encompasses documents concerning the alleged removal of CMI *in connection with retrieval augmented generation (“RAG”)*. RAG is outside of RFP 18’s scope—RAG is a process that occurs *after* a model has been trained.<sup>2</sup>

In response, OpenAI promptly noted that “any alleged removal of CMI via RAG is not within the scope of the RFP as written,” but that OpenAI was “open to meet and confer about the production of this newly requested information—particularly if there is an RFP requesting it.” *Id.* During a call on October 24, OpenAI directly asked The Times to serve a new RFP about CMI removal from RAG to the extent The Times wanted to pursue such discovery. Nikolaides Decl. ¶¶ 5-6. The Times declined to do so and filed this motion instead.

The Times’s motion is an attempt to rewrite an RFP about one aspect of generative AI technology to include documents about an *entirely different aspect* of generative AI technology. Requiring The Times to follow the Federal Rules and serve a new RFP seeking the additional documents it wants would not “unnecessarily inject[] delay into discovery.” Dkt. No. 283 at 3. The Times is not entitled to change the meaning of its discovery requests months after propounding them. *See, e.g., Fung-Schwartz v. Cerner Corp.*, No. 17-cv-0233 (VSB) (BCM), 2020 WL 4927485, at \*8 fn.8 (S.D.N.Y. Aug. 21, 2020) (rejecting an overly broad RFP as written and noting that the court “is not required to edit a party’s discovery request to make it more palatable”); *Patmont Motor Werks, Inc. v. CSK Auto Inc.*, No. 3:04-cv-0473-BES (VPC), 2006 WL 2591042, at \*3 (D. Nev. Sept. 8, 2006) (“The court cannot speculate what may or may not seem obvious to the parties during discovery, and a party responding to discovery requests is entitled to respond to the plain meaning of the request.”) A hypothetical—but unserved—RFP is of no moment. Under Rules 26 and 34, OpenAI must be able to consider, object to, and respond to each RFP on its own terms and on the timeline provided by the Federal Rules. Accordingly, The Times’s motion should be denied.

Sincerely,

MORRISON FOERSTER

KEKER, VAN NEST &  
PETERS LLP

LATHAM & WATKINS  
LLP

/s/ Joseph C. Gratz  
Joseph C. Gratz

/s/ Christopher S. Sun  
Christopher S. Sun

/s/ Elana Nightingale Dawson  
Elana Nightingale Dawson

---

<sup>2</sup> *See* <https://help.openai.com/en/articles/8868588-retrieval-augmented-generation-rag-and-semantic-search-for-gpts> (“RAG is the process of *retrieving* contextual information from a data source and passing that information to a large language model alongside the user’s prompt. This information is used to improve the model’s output (*generated* text or images) by *augmenting* the model’s base knowledge.”) (emphasis in original).